

## Durham Research Online

---

### Deposited in DRO:

28 June 2018

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Bentley, R.A. and Caiado, C.C.S. and Ormerod, P. (2014) 'Effects of memory on spatial heterogeneity in neutrally transmitted culture.', *Evolution and human behavior.*, 35 (4). pp. 257-263.

### Further information on publisher's website:

<https://doi.org/10.1016/j.evolhumbehav.2014.02.001>

### Publisher's copyright statement:

© 2014 This manuscript version is made available under the CC-BY-NC-ND 4.0 license  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

### Additional information:

---

### Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Effects of memory on spatial heterogeneity in neutrally transmitted culture

**R. Alexander Bentley<sup>1</sup>**

*Department of Archaeology and Anthropology, University of Bristol, Bristol, UK*

*r.a.bentley@bristol.ac.uk*

**Camila Caiado**

*Department of Mathematical Sciences, Durham University, Durham, U.K.*

*c.c.d.s.caiado@durham.ac.uk*

**Paul Ormerod**

*Department of Anthropology, Durham University, Durham, U.K.*

*& Volterra Partners LLP, London, U.K.*

*pormerod@volterra.co.uk*

<sup>1</sup>Corresponding author.

## **Abstract**

We explore how cultural heterogeneity evolves without strong selection pressure or environmental differences between groups. Using a neutral transmission model with an isolation-by-distance spatiality, we test the effect of a simple representation of cultural ‘memory’ on the dynamics of heterogeneity. We find that memory magnifies the effect of affinity while decreasing the effect of individual learning on cultural heterogeneity. This indicates that, while the cost of individual learning governs the frequency of individual learning, memory is important in governing its effect.

# Effects of memory on spatial heterogeneity in neutrally transmitted culture

## Abstract

We explore how cultural heterogeneity evolves without strong selection pressure or environmental differences between groups. Using a neutral transmission model with an isolation-by-distance spatiality, we test the effect of a simple representation of cultural ‘memory’ on the dynamics of heterogeneity. We find that memory magnifies the effect of affinity while decreasing the effect of individual learning on cultural heterogeneity. This indicates that, while the cost of individual learning governs the frequency of individual learning, memory is important in governing its effect.

## 1. Introduction

From the foundations of human behavioural ecology (HBE), differences in cultural behavior have been explained as “forms of phenotypic adaptation to varying social and ecological conditions, using the assumption that natural selection has designed organisms to respond to local conditions in fitness-enhancing ways” (Boone and Smith 1998).

Scaled up to group level, the HBE model characterises successful strategies as environment-specific and adaptive, enabling successful groups to out-reproduce competing groups (Henrich et al. 2006). When copying successful behaviours of the community benefits both individual and group, then cooperation can evolve in social networks extending beyond the limits of Hamiltonian inclusive fitness among kin (Henrich et al. 2006; Hill et al. 2011; Hrdy 2009; Rendell et al. 2011).

With new discoveries of cross-cultural variation in behaviours once assumed to be universal (see Nettle 2009a, 2009b), group-selection under different environments has become a more accepted phenomenon in HBE. Cross-cultural variation in the mean offers made in the Ultimatum Game, for example, has been explained through the different benefits of cooperation imposed by different modes of subsistence required in the

environment (Henrich et al. 2006). Norms of attractiveness, for another example, are also related to subsistence (and consequently environment), as humans under low resource conditions tend to be attracted to individuals of larger body mass index (Nettle 2009b). When attractiveness is enhanced by material culture, the stylistic variation may also be adaptive, by maintaining group identity.

The actual details, however, of stylistic traits—such as linguistic dialects, decorative designs and details of folklore (Tehrani and Reide 2009; Evans and Levinson 2009; Kandler and Shennan 2013)—are not specific adaptations to local environment. Within-group cooperation may be advantageous enough in itself—without resort to established models of kin selection, reputation, reciprocity or punishment (Nowak 2006)—such that cooperative norms can evolve among selfish agents modelled only to migrate toward successful communities and copy local strategies (Helbing and Wu 2009; Rand et al. 2009).

As evidenced by cultural phylogenetics (Currie and Mace 2011; Fortunato and Jordan 2009; Tehrani and Reide 2009), stylistic differences between communities in similar environments arise historically, due to chance events and migration over many generations. Even the cross-cultural variation in cooperative norms may be partly explained by demographic differences between groups rather than by their different environments (Lamba and Mace 2011).

To account for historical contingency, over the past decade or so in the context of HBE (Winterhalder and Smith 2000; Nettle et al. 2013), it has become useful to distinguish "evoked culture" from "transmitted culture" (Nettle 2009). Whereas evoked culture is largely determined by environment, transmitted culture is governed by the dynamic equilibrium between social learning and individual learning, as individual learning is disseminated via social learning into evolving cultural traditions (Laland 2004; McElreath & Boyd 2007; Mesoudi 2008).

Crucial to most dynamical models of transmitted culture is the ratio of independent versus social learners in dynamic equilibrium (Rogers 1995; Mesoudi 2008, 2011; Rendell et al.

2011). This mixture can be reduced to a single variable for the fraction,  $\mu$ , of individual learners in the population, and  $(1 - \mu)$  for social learners. Evolutionary theory predicts that social learners  $(1 - \mu)$  can increase in stable environments, and also naturally when the cost of individual learning,  $\mu$ , is high (McElreath & Boyd 2007; Nettle 2009). These approaches assume a selective environment, one where the ‘fitness landscape’ has substantial peaks and individual learners produce the information needed by social learners (“scroungers”) to climb fitness peaks (Mesoudi 2008).

### *Neutral models*

As a logical extreme, “neutral” models can explore cultural evolution on a ‘flat’ fitness landscape, when selective pressures are so weak as to be hypothetically absent. Concerning songbird communication for example, application of a neutral model would assume that in each generation “all subpopulations go through mutation, drift and migration, and all mutant forms are new to the region” (Lynch & Baker 1994: 354). For chaffinches in the Azores, neutral drift within populations was a better explanation than migration for differences between populations (Lynch & Baker 1994). Among warblers of Massachusetts, elements of male-male competition songs were characterised by neutral drift but male-female courtship songs were not, confirming that courtship song elements were selected by the females (Byers et al. 2010).

Comparing tree populations in Panama, Ecuador and Peru, Condit et al. (2005) found that within each region, the similarity (fraction of species shared in common) between small forest plots declined with increasing distance between them, most rapidly at small distances (3-5 km) and then much more slowly further out to 50km, such that similarity decayed linearly with the logarithm of the distance. Condit et al. (2005) found good agreement between these data and Hubbell’s (2001) neutral theory, in which they modeled a landscape of trees which have the same universal probability of death in any time step. When a tree dies, it is either (a) replaced by a copy (descendant) of a randomly-chosen neighbor from a random distance (chosen from a Gaussian distribution), or, with probability  $\mu$ , replaced by a mutant tree of an entirely new species (Condit et al. 2005). This is the neutral model

situated in space, and the process is known as the Moran version because one agent (tree) at a time is selected for replacement.

A substantial insight from such neutral models is that a predominant behavioural norm always emerges through unbiased copying, despite the lack of any fitness difference between the behavioural variants (Neiman 1995). Whatever behaviour emerges as predominant need not be any more adaptive than others, as it can emerge due to different chance histories of individual and social learning. This trend toward predominance is exhibited in the emergent right-skewed distributions of popularity, which closely resemble real data from social and economic contexts (Bentley, Ormerod, Batty 2011; Kandler and Shennan 2013; Ormerod 2012; Reali and Griffiths 2010). This historical contingency means that the same result is unlikely if we were to “replay the tape” of history. The corresponding dynamic turnover in the right-skewed distributions of neutral options (Bentley et al. 2007; Eriksson et al. 2010) is another contrast with optimal adaptations that should not change until the environment changes.

By removing fitness effects, the neutral model allows us to isolate the effects of three components, which we could briefly label as 1) the individual/social learning ratio, 2) distance and 3) memory.

Regarding the *individual/social learning ratio*, neutral approaches typically model successive generations of agents of individual learners  $\mu$  and social learners  $(1 - \mu)$ . In the simplest of these neutral models, individual learning is modelled as random variation, and social learning is modelled through agents randomly sampling behaviours from the previous generation with equal probability (Neiman 1995). More complex versions would impose biases or a network structure (Blythe 2012; Franz & Nunn 2009; Mesoudi and Lycett 2008; Ormerod et al. 2012).

The invention parameter we use is strictly analogous to genetic mutation, but we see the parameter as closely related to the effect of individual learning, in the sense of Boyd and Richerson (1985), through trial and error experimentation, which effectively produces new

variants at specific locations. New variants can also be created, however, through copying errors in the social learning process, so the invention parameter is not exactly a measure of individual learning. Nevertheless, we expect individual learning and ‘invention’ to be strongly correlated, because each new variant qualifies as an invention. Also, to clarify our terms, if a new variant spreads (becomes widespread), the *invention* has become an *innovation* (O’Brien & Shennan 2010; Schumpeter 1934).

To capture *distance* effects generally, we can incorporate assumptions of standard ‘gravity’ models and related ‘isolation by distance’ models. This involves a decay parameter that can be relaxed. Modern hyper-mobility can be translated into these same models when geographic space is transformed into transport network space (Grady *et al.* 2012). Note how this hypermobility contrasts with the trees modelled by Condit *et al.* (2005), whose distribution of mobility is Gaussian and exhibited a linear decline in similarity with the logarithm of distance. We note also that distances also characterise social networks, which can be considered a form of ‘space’, broadly construed as physical, network or even design space.

*Memory* is central to the unique human capacity for goal-directed problem-solving. This is another contrast with most ecological neutral models -- e.g. only living trees are ‘copied’ (Condit *et al.* 2005) – in which there is no ‘memory’ back to lost trees of past generations. Working memory, if simply defined as the “ability to maintain and manipulate thoughts over a brief period” (Wynn and Coolidge 2010: S8), is central to human language use, logic, emotional reasoning, general intelligence, visual and spatial attention, decision making, and planning (Baddley 2001; Wynn and Coolidge 2010). In our model, we consider the effects of cultural memory in a simplified representation, by which spatial location is chosen through neutral decision among many possible locations *and*, subsequently, the choice of behavior is then chosen from among local options. Using the non-spatial neutral model, we previously found that adding memory imposes an ‘egalitarian’ bias on the popularity distribution, making it less right-skewed as memory is increased while holding invention rate constant (Bentley *et al.* 2011). The effects of memory on spatial heterogeneity, however, were not explored. This motivates us to explore how memory

affects, in turn, the effect of the fraction  $\mu$  of individual learners on cultural drift or the strength of isolation-by-distance effects.

Our first hypothesis is that increasing  $\mu$  will increase cultural heterogeneity, by injecting local variation that can be preserved through isolation by distance. Our second hypothesis is that long memory would tend to preserve cultural heterogeneity especially under strong isolation by distance.

## 2. Methods

The model proceeds in a series of repeated iterations. We start with a fixed number of ‘locations’ that could be interpreted as geographical locations, or more generally as social locations. At a given point in time  $t$ , a number,  $n_t$ , of new agents enter the model. Each of these  $n_t$  agents makes two decisions. Firstly, the agent selects a location and secondly it then has to choose amongst the alternative cultural traits available at that location. Agents make each of these choices through random sampling, i.e., with probability proportional to the frequency of the choice among existing agents. In every period, every agent either learns socially from previously available options (with probability  $1 - \mu$ ), or learns individually by inventing something entirely new (with probability  $\mu$ ).

More formally, the algorithm is described by the flowchart in Figure 1. At each time  $t$ , a set number of agents  $n_t$  enter the model. Each agent  $A_{i,t}$ , i.e. the  $i$ -th agent to enter the model at time  $t$  for  $i = 1, \dots, n_t$ , selects a location  $L_{i,t}$  from  $k$  possible locations which follow a multinomial distribution with probabilities proportional to the number of agents in any given location that entered the model in the last  $m$  steps. Once agent  $A_{i,t}$  is assigned to its location  $L_{i,t}$ , it chooses a previously selected cultural trait  $P_{i,t}$  based on preferential attachment or chooses a new trait with probability  $\mu$ , which we call the invention parameter. If the agent chooses to not innovate, its choice is also influenced by its memory  $m$ , i.e. the agent will take into account decisions made by all agents that entered the model in the previous  $m$  time steps, and by the influence of other locations in its own.



We measure the influence of one population into another by a  $k$ -by- $k$  matrix  $W$  where  $W_{i,j}$  describes the influence of the  $i^{\text{th}}$  location in the  $j^{\text{th}}$  location. Therefore, given that the agent chose not to innovate, cultural trait choices follow a multinomial distribution with probability vector given by the proportion of agents that selected each trait in each location within the last  $m$  time steps and reweighted by matrix  $W$  to account for the effect of distance.

---

Figure 1

---

The universal ‘memory’ parameter,  $m$ , which takes integer values from one time step previous to all previous time steps, specifies how much previous history agents take into account in terms of the choices others have made (Bentley et al. 2011). The memory parameter  $m$  specifies that the decisions made by the agents that entered the model in the previous  $m$  steps will influence the decision of new agents. In addition to the choices that an agent has made previously at its own location, the agent can also be influenced by the choices made at the other locations. The importance that the agent assigns to these, relative to the importance of agents at its own location, will vary according to how distant the other locations are. If the number of agents that enter the model at each time step is sufficiently large, we can focus on exploring the impact of memory on the individual learning factors.

The invention parameter  $\mu$  refers to the probability with which a specific agent will deviate from the norm and select a trait that was not previously selected in its own location. This does not necessarily mean that the chosen behaviour is new in the global context of the system, it only means it is new to the local dimension.

Finally, the influence matrix  $W$  assigns weights to the different levels of influence that choices made by agents in other locations might have in the agent’s own location. Here we focus on outcomes when the off-diagonal elements of the influence matrix,  $W$ , are non-

zero. In particular, we are interested in the degree of homogeneity in the aspects of cultural behaviour, which emerges across the different locations. If, for example, agents pay equal weight to trait choices at every location, not just their own, then the outcome will be completely homogeneous, the relative frequency of the various alternative traits will be the same at every location.

Consider now, for example, when the influence of location  $i$  on location  $j$  is assumed to decay exponentially with the square of the distance, as is illustrated in Figure 2.

---

Figure 2

---

The formula for the curves is

$$W_{i,j} = \exp(-\lambda d^2(i,j)) \quad (1)$$

where  $d(i,j)$  is the distance between locations  $i$  and  $j$ .

The special case of  $\lambda = \infty$  has already been explored, as in this case agents only take into account the decisions of agents at their own location (Bentley et al. 2011). This reduces to the non-spatial version of the neutral model, with the diagonal elements of the matrix  $W$  equal to one and all other values zero. This non-spatial neutral model generates an entire family of non-Gaussian, right-skewed popularity distributions, including exponential, power law tails of varying exponents, and power laws across the whole data (Bentley et al. 2010; Evans 2007; Mesoudi and Lycett 2009; Strimling et al. 2009), and also a ‘winner-takes-all’ result when there is no invention at all, i.e.  $\mu = 0$  (Neiman 1995). In addition, the model produces the continuous turnover of rankings of popularity observed empirically within these distributions (Bentley et al. 2007; Eriksson et al. 2010; Evans and Giometto 2011).

Our approach here is to build on these results by exploring finite values of  $\lambda$ . For small values of  $\lambda$ , which we describe as the affinity parameter, the influence of other locations on the choice made by an agent declines rapidly with distance. This distance need not be physical, it could also be a network distance, for example (Grady et al. 2012). We measure the level of homogeneity in the popularity of choices, which emerges across the different locations as follows. We run the model for a 1000 time steps, for a given triplet of values for the memory, invention and affinity parameters, with 1000 agents entering the model at each time step.

In order to assess the level of homogeneity between locations, we calculate the correlations of cultural trait distributions between locations at time  $\tau$  and propose the following measure of similarity

$$S = \frac{1}{k(k-1)} \sum_{1 \leq i < j \leq k} \text{Corr}(P_{L_i, \tau}, P_{L_j, \tau}) \quad (2)$$

where  $k$  is the number of different locations and  $P_{L_i, \tau}$  is a vector that represents the relative proportions of different traits in the  $i^{\text{th}}$  location after  $\tau$  time steps. Each element of this vector represents one of the possible choices, and the vector is long enough to include the maximum possible number of different choices by the end of the run ( $\tau \mu n_t + n_t$ ), which at each location may include zeros for absent choices. The similarity measure  $S$  lies within the interval  $[-1, 1]$ . When  $S$  equals 1, we have total similarity, or in other words global homogeneity. When  $S$  approaches 0, we have maximum heterogeneity. When  $S$  is negative, then the choices in pairwise comparisons tend to be anti-correlated, and as  $S$  approaches -1 they comparisons yield completely contradictory choices between each pair.

In the results that we report here, the locations are placed around a circle. They could equally be placed at random, or in a network. All that we need is a measure of distance between every pair. We repeat the experiment 100 times for the given parameter triplet. Experimentation suggests that this number is more than adequate to assume convergence occurs. We start with  $k = 100$  locations and  $\tau = 1000$  time steps. Concerning the invention fraction, we vary  $\mu$  from 0.005 to 0.05, which is consistent with previous studies that consider mutation rates from  $\mu = 0.001$  to 0.1 (e.g., Lynch and Baker 1994) and similar to

277 ranges proposed for human invention (e.g., Eerkens and Lipo 2005; Diederen et al. 2003;  
278 Srinivasan and Mason 1986; Rogers 1964).

### 280 3. Results

281  
282 To illustrate the character of the results, we first show them for low values of both memory  
283  $m$  and invention fraction  $\mu$  and for single solutions of the model. We therefore set  $m = 1$   
284 and  $\mu = 0.005$ . Figure 3 shows both the weight matrix,  $W$ , and the correlation matrix used  
285 to calculate  $S$  for  $\lambda = 1$ , and Figure 3b sets  $\lambda = 10$ . These are illustrative results from a  
286 single simulation of the model. The locations are placed on a circle, so for location 1, its  
287 nearest neighbours are location 2 on one side, and location 100 on the other.

---

289 Figure 3

---

291  
292 The parameter values  $m = 1$  and  $\mu = 0.005$ , as already noted, generate solutions which  
293 approach ‘winner-takes-all’ when agents at a location only take account of agents at the  
294 same location. Setting  $\lambda = 1$  means that agents assign a high weight to the decisions of a  
295 number of neighbours when choosing from the alternatives available. This means that the  
296 same cultural trait emerges as the ‘winner’ at all locations. There is high correlation  
297 between outcomes at any pair of locations.

298  
299 Figure 4 presents illustrative results for a single solution for the same parameters as in  
300 Figure 3, except with a much stronger decay of influence with distance,  $\lambda = 10$ . The chart  
301 for the weight matrix in Figure 3b, when contrasted with that in Figure 3a, shows that  
302 agents at any given location pay much less attention to decisions made at other locations.

---

304 Figure 4

---

At any given location, the solution is similar to ‘winner-takes-all’, but the trait that wins now differs across the individual locations. This difference is strongest between location pairs coloured blue in Figure 3b (right). Figures 3 and 4 illustrate the impact of varying  $\lambda$ , one of the parameters in the triplet  $(m, \mu, \lambda)$  in a single solution. We now illustrate how varying  $(m, \mu, \lambda)$  impacts the similarity measure  $S$ ; in each case, we fix the memory and invention parameters and simulate the model 100 times for values of  $\lambda$  from 1 to 50.

Figure 5 shows the average of the similarity measure across 100 solutions of the model for given values of  $m$  and  $\mu$  when  $\lambda$  is varied. The top two curves show the results when memory is short, in each case  $m = 1$ , and the bottom two show results with longer memory when  $m = 10$ . There is a strong tendency towards homogeneity across the system when memory is short and the level of homogeneity, or similarity, declines as the rate of invention increases.

These results confirm the illustrative results from a single simulation set out in Figures 3 and 4. For example, the red and the green curves in Figure 5 both have  $\mu = 0.005$ . The curves illustrate quite clearly the importance of memory in the model. For any given value of the affinity parameter, there is considerably more cultural homogeneity when memory is smaller. Higher values of memory mean that, if differences arise in the distribution of cultural traits between locations during the process of solving the model, they have a stronger tendency to persist.

---

Figure 5

---

We also find that the effect of invention on the degree of similarity between locations declines markedly as memory is increased. Consider the green and yellow curves in Figure 5. In each case  $m = 1$ , and  $\mu = 0.005$  and  $0.05$  respectively. The outcomes are clearly different. For small values of the memory parameter, the higher the invention parameter is, the less the overall degree of similarity. Consider now the red and blue curves, where and

$\mu = 0.005$  and  $0.05$  respectively. However, memory is now set at a distinctly higher value,  $m = 10$ . In this case, the impact of varying invention is almost eliminated. With a long memory, the impact of previous choices on an agent's decision about which cultural trait to adopt is higher.

In summary, the similarity measure is impacted by the various parameters in the following ways:

- High values of the affinity parameter, for any given values of invention and memory, mean that an agent assigns low weights to decisions taken by agents in different locations. The higher the affinity parameter,  $\lambda$ , the lower the similarity, in other words the more culturally heterogeneous is the outcome;
- For given values of the affinity parameter, the lower the memory, the higher the degree of similarity, in other words the higher the degree of cultural homogeneity;
- The lower the value of the invention parameter,  $\mu$ , the higher the degree of similarity. However, as memory increases, the effect of varying the invention parameter becomes much less noticeable.

#### **4. Discussion**

In exploring how cultural heterogeneity evolves, memory is important. In smaller societies, collective cultural memory provides a means for humans to situate themselves in their cultural niche and thus cooperate (Pinker 2010). Memory of past observations and encounters allows individuals to anticipate the consequences of future decisions (Giguère and Love 2013; Olivola and Sagara 2009; Stewart et al. 2006). Shared knowledge of other people's feelings helps preserve social relationships amid continual complex negotiations of cooperation (Pinker et al. 2008). Among the !Kung San of the Kalahari Desert, for example, names designate whether two people have a joking relationship or an avoidance relationship (Marshall 1957), which is adaptive for mobile people who may encounter distant relatives infrequently.

We have found that memory is also important in our modelling of cultural heterogeneity under neutral evolution situated in space. Compared to spatial neutral models in an

ecological context without memory (Condit et al. 2005), or cultural neutral models in which the current generation copies from individuals in the previous generation (Neiman 1995; Shennan and Wilkinson 2001; Bentley et al. 2011), it seems appropriate to add memory for cultural phenomena. Cultural variants can be passed between distant generations either directly—when grandma tells a story, so to speak—or when preserved through material culture, written media, or even different cultural groups in which older variants have not yet gone extinct.

Before simulating this spatial neutral model with memory, our hypotheses were that increasing individual learning fraction  $\mu$  would promote spatial heterogeneity, increasing memory  $m$ , and/or strengthening isolation by distance by increasing  $\lambda$ . Although it seemed reasonable at the outset that, under neutral transmission, modelling cultural memory *might* possibly help to preserve local unique inventions and thus increase the effect of individual learning variation on heterogeneity, our modelling shows that, to the contrary, the longer the memory  $m$ , the less likely cultural homogeneity was to emerge. In retrospect it appears that this is because increasing the memory parameter decreases the relative visibility of a new invention, as increasing  $m$  increases the number of choices available to an agent entering the model. Short memory means that cultural traits frequently drop out and become unavailable, because no-one has chosen them in the relevant time frame. With  $m = 1$ , for example, unless a trait has been chosen in the previous time step, it drops out of the system, no matter how many times it has been selected previous to this. With longer memory, however, more traits remain to be selected, and hence the relative size of ‘invention pool’, the number of new alternatives created, becomes very small compared to the number of existing traits.

Regarding individual learning fraction and isolation by distance, our results were more complex than our hypotheses because their effects were not independent of memory. We find that increasing memory  $m$  magnifies the effect of changing the affinity  $\lambda$ , but it decreases the effect of individual learning fraction  $\mu$ . Similar to isolation by distance models, our spatially-aggregated similarity measure decreases as the affinity parameter is increased, but this inverse relationship becomes markedly steeper when memory is

increased. With low memory, for a given value of affinity, increasing invention rate decreases aggregated similarity by introducing random variation. At higher memory values, however, this effect of invention rate vanishes, i.e. even an increase by an order of magnitude in inventiveness has negligible effect under high memory.

## 5. Conclusions

We have shown one basic means by which cultural heterogeneity can evolve under a neutral drift process with memory. This is not in any way meant as a universal explanation, but as substantiation of a powerful alternative hypothesis to selective adaptation to different environments. While it is well established that population size and the individual/social learning ratio are central to neutral evolution, we find that some simple representation of memory increases isolation by distance but decreases the effect of individual learning on cultural heterogeneity. More complex treatments of memory in neutral models, not to mention forward-looking cognitive processes, could underlie new studies of cultural drift contrasting past and present.

The effect of changes in the time scale of this memory is a pertinent evolutionary question, as the Internet paradoxically combines permanent storage of information with shorter attention spans. Over the generational scale, written language accumulates technological knowledge but also regenerates the cultural basis by which people make sense of their own experience, maintain social relationships, and devise scenarios for problem-solving (Carroll 1995; Pinker et al. 2008). It is certainly plausible that information overload is making neutral evolution models more relevant. For the sake of argument, suppose that the Internet makes memory  $m$  longer while decreasing isolation by distance  $\lambda$ . How learning fraction  $\mu$  is changing online is an exciting research question. Other neutral models may incorporate agents with memory, and also with forward expectations (Gureckis and Goldstone 2009). This might be asymmetrically weighed, as experiments suggest people expect less change over the next decade than they report experiencing over the past decade (Quoidbach et al. 2013).



In any case, these changes in memory and individual invention rate brought about by online media surely contrast with millennia of cultural evolution that allowed humans to accumulate information and learn skills over many generations (Henrich 2004; Hruschka et al. 2009; Powell et al 2009). As economist Thomas Schelling put it, most human life consists of individuals responding to a context of other individuals' responses to other individuals. As humans adapt themselves to a 'cognitive niche' of other knowledge-using and cooperative individuals (Pinker 2010), we might consider memory to be the depth of that niche, contemporary population as its length and width, and invention as the ultimate driver of change within it.

## References

- Baddeley, A. (2001). Is working memory working? *American Psychologist* 56, 851–864.
- Bentley, R.A., P Ormerod & Batty, M. (2011). Evolving social influence in large populations. *Behavioral Ecology and Sociobiology* 65: 537–546.
- Bentley, R.A., Lipo, C.P., Herzog, H.A. & Hahn, M.W. (2007). Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior* 28, 151–158.
- Blythe, R.A. (2012) Random copying in space. *Advances in Complex Systems* 15, 1150012.
- Boone, J. L. & Smith, E.A. (1998) Is it evolution yet? A critique of evolutionary archaeology. *Current Anthropology* 39, S141–S173.
- Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. The University of Chicago Press, London.
- Byers, B.E., Belinsky K.L. & Bentley, R.A. (2010). Independent cultural evolution of two song traditions in the chestnut-sided warbler. *American Naturalist* 176, 476–489.
- Carrol, J. (1995). Evolution and literary theory. *Human Nature* 6:119–134.
- Condit, R., Pitman, N, Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E., & Hubbell, S.P. (2005). Beta-diversity in tropical forest trees. *Science* 295: 666–669.
- Currie, T.E. and Mace, R. (2011) Mode and tempo in the evolution of socio-political organization. *Phil. Trans. R. Soc. B*, 366: 1108–1117
- Diederer, P., van Meijl, H. & Wolters, A. (2003). Modernisation in agriculture: What makes a farmer adopt an innovation? *International Journal of Agricultural Resources, Governance and Ecology* 2, 328–342.

461 Eerkens, J. W. & Lipo, C.P. (2005). Cultural transmission, copying errors, and the generation of  
462 variation in material culture and the archaeological record. *Journal of Anthropological*  
463 *Archaeology*, 24, 316–334.

464 Eriksson, K., Jansson, F. & Sjöstrnad, J. (2010). Bentley’s conjecture on popularity toplist turnover  
465 under random copying. *The Ramanujan Journal*, 23, 371–396.

466 Evans, T.S. (2007). Exact solutions for network rewiring models. *European Physics Journal B* 56:  
467 65–69.

468 Evans, T. & Giometto, A. (2011), Turnover rate of popularity charts in neutral models.  
469 arXiv:1105.4044 [physic.soc-ph].

470 Evans, N. & S.C. Levinson (2009). The myth of language universals: Language diversity and its  
471 importance for cognitive science. *Behavioral and Brain Sciences* 32: 429–492.

472 Fortunato, L. & Jordan, F.M.(2010) Your place or mine? A phylogenetic comparative analysis of  
473 marital residence in Indo-European and Austronesian societies. *Philosophical Transactions*  
474 *of the Royal Society B* 365: 3913–3922.

475 Franz, M. & Nunn, C. L. (2009). Rapid evolution of social learning. *Journal of Evolutionary*  
476 *Biology* 22, 1914–1922.

477 Grady, D., Thiemann, C. & Brockmann, D. (2012) Robust classification of salient links in complex  
478 networks. *Nature Communications* 3, Article 864

479 Gureckis, T.M. & Goldstone, R.L. (2009). How you named your child: Understanding the  
480 relationship between individual decision making and collective outcomes. *Topics in*  
481 *Cognitive Science* 1, 651–674.

482 Giguère G. & Love B.C. (2013) Limits in decision making arise from limits in memory retrieval.  
483 *Proceedings of the National Academy of Sciences USA*, in press

484 Helbing, D. & Yu, W. (2009) The outbreak of cooperation among success-driven individuals under  
485 noisy conditions. *Proceedings of the National Academy of Sciences* 106:3680–85.

486 Henrich, J. (2004). Demography and cultural evolution: Why adaptive cultural processes produced  
487 maladaptive losses in Tasmania. *American Antiquity*, 69, 197–214.

488 Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C., McElreath, R., Gurven, M.,  
489 Hill, K., Barr, A., Ensminger, J., Tracer, D., Marlow, F., Patton, J., Alvard, M., Gil-White F.  
490 & Henrich, N. (2005) ‘Economic Man’ in cross-cultural perspective: Ethnography and  
491 experiments from 15 small-scale societies. *Behavioral and Brain Sciences* 28:795–855.

492 Hill, K.R., Walker, R.S., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A.M.,  
493 Marlowe, F., Wiessner, P. & Wood, B. (2011). Co-residence patterns in hunter–gatherer  
494 societies show unique human social structure. *Science* 331, 1286–1289.

495 Hrdy, S.B. (2009). *Mothers and Others: The Evolutionary Origins of Mutual Understanding*.  
 496 Cambridge, MA: Belknap Press.

497 Hruschka, D.J., Christiansen, M.H., Blythe, R.A., Croft, W., Heggarty, P., Mufwene, S.S.,  
 498 Pierrehumbert J.B & Poplack, S. (2009). Building social cognitive models of language  
 499 change. *Trends in Cognitive Sciences* 13(11),464-469.

500 Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton  
 501 University Press.

502 Kandler, A. & Shennan, S.J. (2013). A non-equilibrium neutral model for analysing cultural  
 503 change. *Journal of Theoretical Biology*, in press.

504 Laland, K.M. (2004). Social learning strategies. *Learning & Behavior*, 32, 4–14.

505 Lamba, S. & Mace, R. (2011). Demography and ecology drive variation in cooperation across  
 506 human populations. *Proceedings of the National Academy of Sciences USA* 108(35):14426–  
 507 14430.

508 Lynch, A. & Baker, A.J. (1994). A population memetics approach to cultural evolution in chaffinch  
 509 song. *Evolution* 48(2),351-359

510 McElreath, R. and Boyd, R. (2007). *Mathematical Models of Social Evolution: a Guide for the*  
 511 *Perplexed*. Chicago: University Press.

512 Marshall, L. (1957). The kin terminology of the !Kung Bushmen. *Africa* 27, 1 – 25.

513 Mesoudi, A. (2008). An experimental simulation of the “copy-successful-individuals” cultural  
 514 learning strategy: adaptive landscapes, producer–scrounger dynamics, and informational  
 515 access costs. *Evolution and Human Behavior* 29, 350–363.

516 Mesoudi, A. (2009). How cultural evolutionary theory can inform social psychology and vice versa.  
 517 *Psychological Review*, 116(4):929–952.

518 Mesoudi, A. & Lycett, S.J. (2009). Random copying, frequency-dependent copying and culture  
 519 change. *Evolution and Human Behavior* 30: 41–48.

520 Mesoudi, A. (2011). An experimental comparison of human social learning strategies: payoff-  
 521 biased social learning is adaptive but underused. *Evolution and Human Behavior* 32, 334–  
 522 342.

523 Neiman, F.D. (1995). Stylistic variation in evolutionary perspective. *American Antiquity*, 60, 7–36.

524 Nettle, D. (2009a). Ecological influences on human behavioural diversity: a review of recent  
 525 findings. *Trends in Ecology and Evolution*, 24(11):618–624.

526 Nettle, D. (2009b). Beyond nature versus culture: cultural variation as an evolved characteristic.  
 527 *Journal of the Royal Anthropological Institute*, 15:223–240.

528 Nettle, D., Gibson, M.A., Lawson, D.W. & Sear, R. (2013) Human behavioral ecology: current  
529 research and future prospects. *Behavioral Ecology*, in press.

530 Nowak, M.A. (2006) Five rules for the evolution of cooperation. *Science* 314: 1560--1563.

531 Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M.A. (2006). A simple rule for the evolution of  
532 cooperation on graphs and social networks. *Nature* 441:502--505.

533 Olivola, C.Y. & Sagara, N. (2009) Distributions of observed death tolls govern sensitivity to human  
534 fatalities. *Proceedings of the National Academy of Sciences USA*, 106, 22151--22156.

535 Ormerod, P. (2012) *Positive Linking: How Networks Can Revolutionise the World*. London: Faber  
536 and Faber.

537 Pinker, S. (2010) The cognitive niche: Coevolution of intelligence, sociality, and language.  
538 *Proceedings of the National Academy of Sciences USA* 107, 8993--8999.

539 Pinker, P., Nowak, M.A. & Lee, J.J. (2008) The logic of indirect speech. *Proceedings of the*  
540 *National Academy of Sciences USA* 105, 833--838.

541 Powell, A., Shennan, S. & Thomas, M.G. (2009). Late Pleistocene demography and the appearance  
542 of modern human behavior. *Science* 324, 1298--1301.

543 Quoidbach J., Gilbert, D.T. & Wilson, T.D. (2013) The end of history illusion. *Science* 339:96--98.

544 Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. (2009). Positive interactions  
545 promote public cooperation. *Science* 325, 1272--1275.

546 Reali, F. & Griffiths, T. L. (2010) Words as alleles: Connecting language evolution with Bayesian  
547 learners to models of genetic drift. *Proceedings of the Royal Society B* 277:429--36.

548 Rendell, L., Fogarty, L., Hoppitt, W.J.E., Morgan, T.J.K., Webster, M.M & Laland, K.N. (2011).  
549 Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in*  
550 *Cognitive Sciences*, 15: 68--76.

551 Rogers, E.M. (1964). *Diffusion of Innovations*. Glencoe, IL: Free Press.

552 Schumpeter, J. A. (1934) *The Theory of Economic Development*, Harvard University Press.

553 Shennan, S.J. & Wilkinson, J.R. (2001). Ceramic style change and neutral evolution. *American*  
554 *Antiquity* 66:577--594.

555 Srinivasan, V. & Mason, C.H. (1986). Nonlinear least squares estimation of new product diffusion  
556 models. *Marketing Science* 5, 169--178.

557 Stewart, N., Chater, N. & Brown, G.D.A. (2006) Decision by sampling. *Cognitive Psychology*,  
558 53:1--26.

559 Strimling, P., Sjöstrand, J., Enquist, M. Eriksson, K. (2009). Accumulation of independent cultural  
560 traits. *Theoretical Population Biology* 76: 77--83.

561 O'Brien, M. J. & Shennan, S. J. (2010) Issues in anthropological studies of innovation. In:

562        *Innovation in cultural systems: Contributions from evolutionary anthropology*, ed. M. J.  
 563        O'Brien and S. J. Shennan, pp. 3-17, MIT Press.  
 564        Ormerod, P., Tarbush, B. & Bentley, R.A. (2012). Social network markets: the influence of network  
 565        structure when consumers face decisions over many similar choices.        Cornell  
 566        University arXiv:1210.1646  
 567        Tehrani, J. J. & Riede, F. (2008). Towards an archaeology of pedagogy: learning, teaching and the  
 568        generation of material culture traditions. *World Archaeology*, 40(3):316–331.  
 569        Winterhalder, B. & Smith, E.A. (2000). Analyzing adaptive strategies: Human behavioral ecology  
 570        at twenty-five. *Evolutionary Anthropology* 9, 51–72.  
 571        Wynn, T. & Coolidge, F.L. (2010). Beyond symbolism and language. *Current Anthropology* 51,  
 572        S5-S16.  
 573

## Figure legends

**Figure 1.** Flowchart representing the algorithm described in Section 2. Here we show the solution for any weight matrix  $W$ . In the examples described throughout this paper, we use  $W$  as described in Equation 1.

**Figure 2.** For a given location, the weight assigned by an agent to the choices made in other locations. Distance on the x-axis and the y-axis shows the weight.

**Figure 3.** Weak decay by distance ( $\lambda = 1$ ). **Left:** weight matrix,  $W$ , for  $\lambda = 1$ ,  $m = 1$  and  $\mu = 0.005$ . The axes show the location number (note that locations are in a circle, so location 100 is adjacent to location 1). The colour codes show the weight associated between each location pair. **Right:** the correlation matrix between the cultural trait distributions in each location. Note all correlations between location pairs are high and the calculated  $S$  measure for this simulation is 0.99.

**Figure 4.** Strong decay by distance ( $\lambda = 10$ ). **Left:** the weight matrix,  $W$ , for  $\lambda = 10$ ,  $m = 1$  and  $\mu = 0.005$ . The axes show the location number. The colour codes show the weight associated between each location pair. **Right:** the correlation matrix between the cultural trait distributions in each location. The calculated  $S$  measure for this simulation is 0.10.

**Figure 5.** Average of the similarity  $S$ , as a function of influence decay parameter  $\lambda$ , across 100 solutions of the model for several combinations of  $m$  and  $\mu$ . Red:  $m = 10$ ,  $\mu = 0.005$ ; Blue :  $m = 10$ ,  $\mu = 0.05$ ; Yellow: :  $m = 1$ ,  $\mu = 0.05$ ; Green: :  $m = 1$ ,  $\mu = 0.005$ .











